Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.017

IJCSMC, Vol. 7, Issue. 5, May 2018, pg.154 – 159

Providing Novel Approach for Classification of Soil and Crop Prediction Using Decision Support System

Vrushal Milan Dolas[1], Prof. Uday Joshi[2]
[1]Department of Computer Engineering, K. J. Somaiya College of Engineering, Mumbai, India
[2]Department of Computer Engineering, K. J. Somaiya College of Engineering, Mumbai, India
[1] vrushal.dolas20@gmail.com

Abstract— Decision Support System has been applied to solve variety of agricultural problems. It provides the framework that allows both the decision-makers and farmers to make good decisions. Decision support systems (DSS) use databases, human– machine to combine a large number of models to realize scientific decision-making. The environmental data plays a vital role in agriculture decision, which changes at a rapid rate. Keeping these data updated can be done by using a Service Oriented Approach (SOA). The goal of the irrigation system is to determine the exact amount of needed water and the exact timing for applying it also the soil type required for particular crop. The amount of water applied is determined dependent on each user situation. The system can use to build a decision-making model to for irrigation for different users. Decision tree is a well known approach for classification in data mining. C4.5 and Classification and Regression Trees (CART) are two widely used decision tree algorithms for classification. The main drawback of C4.5 algorithm is that, it is biased towards attributes with more values while CART algorithm produces misclassification errors when the domain of the target attribute is very large. In view of these limitations, this paper presents a modified decision tree algorithm. The test proves that the modified decision tree algorithm has higher classification accuracy when compared to C4.5 and CART algorithms.
Keywords— Decision Support System, Service Oriented Approach, CART algorithm, decision-making model.

I. INTRODUCTION

Agriculture is back bone business in India. It contributes 10-15% GDP to the India economy. In Indian agriculture, the volume of data is enormous. The data when become information is highly useful for many purposes. The conventional and traditional system of data analysis in agriculture is purely dependent on statistics. Data mining is a modern data analysis technique. It has wide range of applications in the field of agriculture. In this work, applications of the data mining techniques in the area of agriculture and its allied areas are studied. Different techniques of data mining have been used in this field.

Data Mining has emerged as one of the major research domain in the recent decades in order to extract implicit and useful knowledge. Classification techniques in data mining are capable of processing a large amount of data. Classification of data is one of the major steps towards extracting useful information in data

mining. Plantation of crop according to the soil type is very important in the success of crop. There are many characteristics which decide the nature of the soil ex: pH (power of Hydrogen) value, moisture, Ec (Electrical Conductivity), ESP (Exchangeable Sodium Percentage), etc. The pH value is used to decide whether the soil is acidic, the Ec value is used to decide whether the soil is saline (salt content) and the ESP value is used to decide whether the soil is alkali. The soil samples that belong to either acidic or saline or alkali are supposed to be problematic soils as they are not conducive for crop growth. The other soil samples are non-problematic as they are conducive for crop growth. To analyze the soil type of a geographical area, soil samples are collected then the samples are classified into different types. Using data mining techniques one can efficiently classify the soil samples into different categories.

A. Motivation

The purpose of this work is to categorize the soil samples according to the nutrients present into it and to predict the crops which can be yield in the particular soil. Along with focus on the soil classification based on decision tree this work focuses on prediction of need of exact amount of water to apply to the field and the exact timing for application.

B. Scope

- Classification of soil using a modified decision tree approach
- Prediction of crops for suitable land.
- Suggestions for improving the quality of land using various nutrients.
- Amount of water to be supplied to crop with respect to climatic conditions using Decision support system (DSS).
- Suggests suitable crop for a particular type of soil.

II. Literature Survey

The literature for the classification of the soil and prediction of crops are as follows-

In 2014, K AdityaShastry, Sanjay H A and Kavya H discussed various classification techniques in the paper "A Novel Data Mining Approach for Soil Classification". The algorithms CART, C4.5 and proposed approach has been studied. Accuracy obtained for C4.5 is more as compared to CART. The proposed system uses gini index and gini ratio. The impurity measure gini index used in CART is biased towards attributes with higher range of values, while information gain used in C4.5 is biased towards attributes with high values. This drawback is removed using the proposed approach. From the paper it is also observed that the manual soil classification being done is very cumbersome and time consuming.

In 2015, Monali Paul, Santosh K. Vishwakarma and Ashok Verma in "Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach" introduced that Data mining in agriculture field is somewhat a novel research field. Data mining is the process of discovering unknown and likely impressive patterns in large datasets. Steps of data mining such as selection, preprocessing, transformation, data mining and interpretation has been discussed and Naïve bayes and K-nearest neighbor has been used for prediction and analysis. Naive Bayes classifiers can be trained very efficiently in a supervised learning setting and works much better with complex real situations. K-Nearest Neighbor makes predictions based on the outcome of the K neighbors closest to that point. K- nearest Neighbour uses Euclidean distance formula for calculation.

In "Sensible approach for Soil Fertility Management using GIS Cloud" Decision Support System (DSS) using GIS enabled cloud technologies is implemented. The proposed approach of agricultural information development and integration system ensures that complete agricultural related data on cloud database can be integrated into spatial maps through GIS technologies, to organize, accumulate, and administer geospatial data in a cloud database according to individual farmer land information for improving data accuracy of digital agriculture fertilizer management. IaaS, PaaS and Saas has been discussed with the respective layers used. GIS Cloud server is used to incorporate, integrate, store, update and manage complete information of agriculture.

In 2015, Amol D. Vibhute, K. V. Kale, Rajesh K. Dhumal and S. C. Mehrotra in "Soil Type Classification and Mapping using Hyperspectral Remote Sensing Data" has demonstrated use of support vector machine algorithm for identification, mapping and classification of various types of soil using high spectral resolution Hyperspectral data. Gaussian Radial Basis Function (RBF) kernel of SVM was used and the accuracy obtained is 71.18.

In 2010, Sofianita Mutalib, S-N-FadhlunJamian, Shuzlina Abdul-Rahman and Azlinah Mohamed explained In "Soil Classification: An Application of Self Organizing Map and k-means" The two unsupervised technique Kohonen Self Organizing Map (SOM) and k-means have been used to classify the soil. Characteristics of soil such as parent material, soil horizontal profile, color of soil, texture of soil and soil depth is listed. This paper predicts the classification of soil and gives information about the plants to be cultivated in specify type of soil.

In 2014, Ayman Nada, Mona Nasr, MarwaSalah[6],have proposed a decision support system for agriculture. The goal of the work described here is to optimize crop water usage by using SOA technology. Their propose task aims to compute the proposed irrigation schedule. The needed water in each plant stage is calculated according to the user specific situation which determined in the previous subtask. This paper presents that SOA used to get precise irrigation schedule for mango trees.

In 2014, Ayman Nada, Mona Nasr, MarwaSalah, have proposed a decision support system for agriculture. The goal of the work described here is to optimize crop water usage by using SOA technology. Their propose task aims to compute the proposed irrigation schedule. The needed water in each plant stage is calculated according to the user specific situation which determined in the previous subtask. This paper presents that SOA used to get precise irrigation schedule for mango trees.

III. PROPOSED SYSTEM

The primary objective of this system is to classify the soil according to the nutrients into it. For this, we have taken datasets of soil samples. The soil will be classified using decision tree algorithm and type of soil will be displayed. Also we are going to predict the crops suitable for the particular type of soil. In addition to this, we are going to improve the soil if the farmer wants to yield particular crop in the same soil by suggesting the requirements of the nutrients for the same soil. Along with focus on the soil classification based on decision tree this work focuses on prediction of need of exact amount of water to apply to the field and the exact timing for application using Decision support system (DSS).

C4.5 is an algorithm to generate a decision tree which uses an impurity measure called information gain which is the difference in entropy. Classification And Regression Trees (CART) is a classification method which uses historical data to construct decision trees. It is based on binary recursive partitioning where the parent nodes are always split into exactly two nodes and the process is repeated by treating each node as a parent. It uses Gini index as the impurity measure for splitting.

The main drawback of C4.5 algorithm is that it produces large decision trees with more misclassification errors. This is due to the fact that it uses information gain as the splitting measure which is biased towards attributes with more values. The main drawback of CART algorithm is that it does not use combinations of variables. Also, it may result in unstable trees where a change in the sample may give different trees. This is due to the fact that gini index is used as the splitting measure which encounters problems when the domain of the target attribute is relatively wide.

To overcome these drawbacks we propose a modified decision tree approach for soil classification. Here, we first calculate the gini index for different ranges of attribute values instead of computing for every successive pairs as was done in CART algorithm. Then we have used ratios of these computed gini indices to reduce the bias that was introduced by information gain in C4.5 algorithm.

The proposed algorithm is described below.

Input: Experimental data set D

Output: A decision tree T which is created by giving experimental dataset.

1) Construct a root node which contains the whole data set

2) Compute Gini index for different ranges of attribute values of pH, Ec and ESP

3) Compute the ratio of the gini indices for each attribute

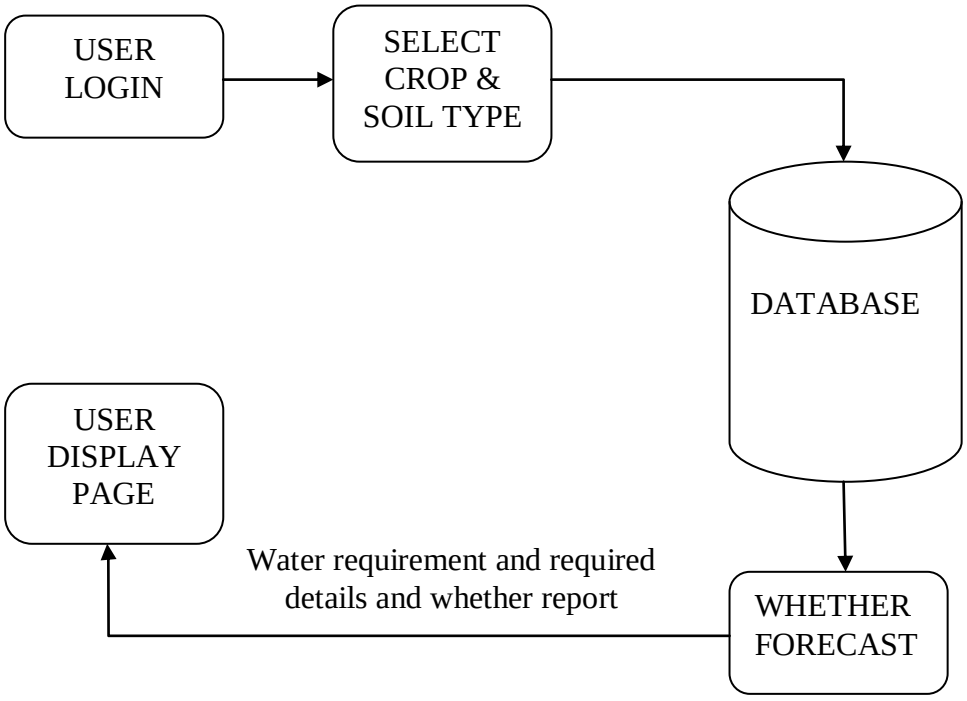
4) Select an attribute with the least gini ratio as the node to be split.

5) Split the examples of the current node into different subsets based on values of the selected attributes

6) Create a new node as a child of the current node for each subset and pass the examples in the subset to the node

7) Recursively repeat steps 2 to 5 until further splitting is not possible (only one instance remains in the node).

IV. ARCHITECTURE OF PROPOSED METHODOLOGY



Step wise execution:

Step 1: user/farmer can registered to the application using user registration link.

Step 2:After login, user/farmer can edit their own profile.

Step 3:For Water requirement calculation user have to select crop from database.

Step 4:Then the result shows the current temperature, soil type required for crop, required season & water required for that crop in user interface by using knowledge based & database management decision support system.

Step 5: User can check the previous year's rainfall record by using database system used in the DSS.

V. COMPARISON BETWEEN EXISTING SYSTEM AND PROPOSED SYSTEM

EXISTING SYSTEM:

The existing work is focused only on the development of a new soil classifier based on decision tree. Its main aim was to develop a soil classifier which would classify soils into acidic, alkali, saline and non-problematic soils with a higher accuracy. There is no efficient prediction method for agricultural data like crops using efficient prediction methods and compare it with other prediction methods like regression, etc. in terms of accuracy, size of the resulting model and computational effort.

Proposed system:

In proposed system along with focus on the soil classification based on decision tree this work focuses on prediction of need of exact amount of water to apply to the field and the exact timing for application using Decision support system (DSS). Also it ssuggests suitable crop for a particular type of soil.

VI. CONCLUSION

Classifying the soil using proposed algorithm helps to know the overall health of the soil and the content of the nutrients in the soil. Also prediction helps to decide what all he can yield in particular soil. This can help the farmer to yield the variety of crops seasonally. In case we wish to yield particular crop, then we can improve the soil by adding the necessary nutrients in the soil as per required by that crop. The amount of water required is determined dependent on each user situation. This system is plays a role as expert in the agriculture field. Along with the water requirement calculation, this system can also help farmers in problems related to agriculture. The system can be used to build a decision-making model for irrigation for different users. This is focused only on services and the system assembly model, which leads to a good reflection, wide usability and save time.

VII. FUTURE WORK

Modern agriculture offer a range of benefits including greater production and higher income for farmers in both developed and developing countries for e.g. using a web application for take a decision in water requirement calculation and other problems regarding to the agriculture so that increase a production of field in available environmental situation. Farmer empowerment can be successfully achieved by providing them right information at right time. Informed decision making by effectively utilizing the different new technology and different agriculture model tools will increase productivity. This will not only motivate the farmers in the rural area, but also drive urban youth towards hi-tech farming.

ACKNOWLEDGEMENT

I would like to express my gratitude and appreciation to all those who helped me to complete this paper. A special thanks to my guide, **Prof. Uday Joshi** whose help, stimulating suggestions and encouragement helped me in writing this paper.

REFERENCES

[1] K AdityaShastry, Sanjay H A and Kavya H "A Novel Data Mining Approach for Soil Classification" The 9th International Conference on Computer Science & Education (ICCSE 2014) August 22-24, 2014. Vancouver, Canada.

[2] Monali Paul, Santosh K. Vishwakarma and Ashok Verma "Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach" 2015 International Conference on Computational Intelligence and Communication Networks.

[3] Amol D. Vibhute, K. V. Kale, Rajesh K. Dhumal and S. C. Mehrotra "Soil Type Classification and Mapping using Hyperspectral Remote Sensing Data" 2015 International Conference on Man and Machine Interfacing (MAMI).

[4] SofianitaMutalib, S-N-FadhlunJamian, Shuzlina Abdul-Rahman and Azlinah Mohamed "Soil Classification: An Application of Self Organising Map and k-means" 2010 10th International Conference on Intelligent Systems Design and Applications.

[5] www.en.wikipedia.org

[6] Ayman Nada, Mona Nasr, Marwa Salah "Service oriented approach for decision support systems" Information Technology and Artificial Intelligence Conference (ITAIC), 2014 IEEE 7th Joint International

